

Ingeniería Inversa Aplicada a Software Numérico: Modelos Climáticos

Fernando G. Tinetti*, Pedro G. Cajaraville[#], Juan C. Labraga^{##}, Mónica A. López^{##}, María G. Olguín[#]

[#]Departamento de Informática, Facultad de Ingeniería – (UNPSJB¹)

^{##}Centro Nacional Patagónico (CENPAT-CONICET²)

III-LIDI, Facultad de Informática – (UNLP³)

fernando@info.unlp.edu.ar, gustavo.cajaraville@gmail.com, {labraga, monica}@cenpat.edu.ar, gabriela.olguin@gmail.com

RESUMEN

En este proyecto se aplicarán las estrategias y métodos de ingeniería inversa al software que implementa un modelo numérico del clima. El producto final es una documentación de la estructura general y la forma en que se comunican los modelos involucrados. Más específicamente, en principio este trabajo incluirá la documentación de las rutinas de mayor nivel de abstracción, hasta el segundo nivel del árbol de invocación. El modelo climático (numérico) es en realidad la composición o acoplamiento de dos modelos casi independientes: el modelo atmosférico y el modelo oceánico. Justamente se parte de lo que se denomina el *modelo acoplado*, que consta del programa principal *encargado* de invocar a rutinas de ambos modelos: atmosférico y oceánico (primer nivel de rutinas). En el segundo nivel de rutinas se encuentran las invocaciones directas desde las rutinas de estos dos últimos modelos. Además del propio programa en código fuente Fortran 77 lamentablemente se cuenta con muy poca documentación, la mayor parte de la cual corresponde más al área numérica de los procesos físicos del clima que de ingeniería de software *aplicada o usada* para el desarrollo de la aplicación. En este sentido, es bastante usual en el campo de los modelos numéricos el punto de partida: software heredado (*legacy code*) en la forma de un programa de varios miles de líneas que se considera la implementación directa de un modelo numérico. Esto por supuesto dificulta las tareas de mantenimiento, cambios de funcionalidades y aún la utilización misma de la aplicación, todas tareas que se intentan simplificar con la aplicación de ingeniería inversa para recuperar información de desarrollo/ingeniería de software.

Palabras Clave: ingeniería inversa, software heredado, modelado climático.

1. INTRODUCCIÓN

La comprensión del sistema climático es un problema de gran interés científico mundial. Si bien es cierto que se han realizado avances considerables en el tema, aún son muchos los factores que continúan limitando la capacidad de detectar, atribuir y comprender el cambio climático actual y proyectar los cambios climáticos que podrían ocurrir en el futuro [1]. En la actualidad, las herramientas más confiables para la investigación del clima, sus fluctuaciones y variaciones, son los Modelos de Circulación General de la Atmósfera (MCGA).

¹ Universidad Nacional de la Patagonia San Juan Bosco

² Centro Nacional Patagónico - Consejo Nacional de Investigaciones Científicas y Técnicas

³ Universidad Nacional de La Plata

* Investigador Asistente, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

Un MCGA es una representación numérica espacial y temporal aproximada de los principales procesos físicos que ocurren en la atmósfera y de las interacciones con los otros componentes del medio ambiente. Esencialmente, está constituido por un sistema de ecuaciones diferenciales en derivadas parciales que expresan los principios de conservación de la cantidad de movimiento, la energía termodinámica y la masa del sistema. Estas ecuaciones expresan las leyes físicas que gobiernan la atmósfera. Actualmente, los MCGA están integrados con modelos numéricos de funcionamiento de los océanos, la criosfera, y representaciones simplificadas de la biosfera (MCGA *acoplados*), lo que permite simular las interacciones de la atmósfera con su entorno, en distintas escalas de tiempo.

Mediante el uso de computadoras con gran capacidad de procesamiento y la aplicación de diversos métodos de cálculo numérico, pueden obtenerse soluciones numéricas aproximadas del sistema de ecuaciones de un MCGA. De este modo, se obtiene la evolución temporal y espacial (en forma tridimensional) del sistema climático, en función de las condiciones iniciales y de contorno elegidas y de los valores de ciertos parámetros climáticos (ejemplo: concentración de CO₂ atmosférico). La solución numérica de un MCGA en condiciones preestablecidas se denomina *experimento climático*.

1.1 El Modelo CSIRO (Commonwealth Scientific and Industrial Research Organisation)

El modelo de circulación general de la atmósfera CSIRO versión MII, es uno de los MCGA utilizado por la comunidad científica internacional. Fue desarrollado en la División de Investigaciones Atmosféricas, CSIRO, Australia [2] y participó junto con otros modelos del *Program for Climate Model Diagnosis and Intercomparison* (PCMDI) [9] [3] [4] [5].

En el año 1999 fue cedido, luego de un período de capacitación, a un grupo de investigación de la Unidad de Investigación de Oceanografía Física y Meteorología (UIOM) del CENPAT (Centro Nacional Patagónico)-CONICET. Las aplicaciones del modelo CSIRO MII en la UIOM han permitido avances en el conocimiento de temas cruciales para el desarrollo económico y social de la Argentina: la predictibilidad de anomalías climáticas estacionales, el impacto del calentamiento global sobre el ciclo hidrológico en la, el calentamiento global y los cambios climáticos regionales y el impacto del cambio climático en los recursos hídricos de la cordillera de los Andes. El sistema informático de pronóstico de anomalías climáticas basado en el modelo CSIRO MII desarrollado por el equipo de la UIOM, publica mensualmente en forma experimental información objetiva sobre la magnitud y la probabilidad de condiciones atípicas del clima.

1.2 Ingeniería Inversa

La ingeniería inversa del software es el proceso consistente en analizar un programa, en un esfuerzo por crear una representación del mismo con un nivel de abstracción más elevado que el código fuente. Es un proceso de recuperación del diseño. Las herramientas a utilizar extraen información acerca de los datos, arquitectura y diseño de procedimientos de un programa ya existente. A diferencia de lo usual en otras disciplinas, donde este proceso se aplica a los productos de la competencia, en ingeniería de software con frecuencia se debe aplicar a los propios trabajos de la organización, realizados hace muchos años. En el caso de este proyecto, se aplica a un trabajo hecho por otra organización, pero que se intenta optimizar con varios objetivos: desde la optimización de rendimiento numérico hasta el cambio/mejora de funcionalidades.

En el proceso de recuperación del diseño, además de las observaciones del sistema se añade conocimiento del dominio de la aplicación, información externa y procesos deductivos. Esto

significa recoger más información de la que se puede obtener del código fuente. Se ponen en juego, además de la experiencia que se haya desarrollado con el sistema, los conocimientos generales sobre los dominios del problema y de la aplicación. Los objetivos de la ingeniería inversa, algunos de los cuales se intentan mantener en este proyecto son, de menor a mayor nivel de abstracción:

- Representaciones de diseño de procedimiento.
- Información de las estructuras de datos.
- Modelos de flujo de datos y de control.
- Modelos de entidades y de relaciones.

Según va aumentando la abstracción va aumentando la complejidad del trabajo, así como la necesidad de comprensión de la aplicación. La ingeniería inversa debe ser capaz de abstraer, a partir del código fuente, información significativa del procesamiento que se realiza, las estructuras de datos que se usan en el programa y el interfaz con el usuario que se utiliza.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Para la comprensión de la estructura del programa o programas de cómputo secuencial que implementan el modelo acoplado océano-atmósfera, se utilizará un método de refinamiento sucesivo [8]. En principio se identificarán bloques de rutinas y sus relaciones funcionales, para luego analizar más en detalle las rutinas más significativas, identificando entradas, salidas y procesos involucrados, y redocumentándolas. La idea subyacente, es aprovechar los objetivos de la ingeniería inversa para:

- *Comprender el procesamiento*: se utiliza un método de refinamiento de mayor a menor nivel de abstracción. Se establecen los siguientes pasos:
 1. Diagrama de bloques de los programas que componen el sistema y sus relaciones funcionales.
 2. Diagrama de módulos para cada uno de los programas. En este punto se estudia el interior de cada módulo. En general se definen las siguientes funcionalidades:
 - 2.1. Preprocesamiento (se preparan los datos para el punto siguiente).
 - 2.2. Procesamiento.
 - 2.3. Postprocesamiento (se preparan los datos para exportarlos desde este módulo).
- *Comprender los datos*: se divide, a su vez en dos partes:
 1. Análisis de los datos internos.
 2. Estudio de las estructuras globales de datos (archivos, bases de datos, etc.).
- *Representar interfaces de usuario*: en [8] se menciona el álgebra de procesos que permite representar el interfaz en base a agentes y acciones. Básicamente consiste en una notación taquigráfica que permite obtener una idea completa del interfaz. Estas tareas no se llevarán a cabo al menos a corto plazo en el desarrollo de este proyecto.

Más específicamente en el contexto de este proyecto, las tareas de investigación inmediatas se llevarán a cabo sobre el modelo CSIRO MII y están directamente relacionadas con el estudio del modelo numérico desde el punto de vista del procesamiento secuencial. Esto implica, entre otras tareas:

- La comprensión del modelo numérico de base para el modelo computacional acoplado océano-atmósfera. Se deben estudiar los procesos modelados numéricamente. [1] [11] y luego, de manera más específica, los modelos oceánico y atmosférico por separado. Una vez que se han analizado y/o estudiado estos modelos, se debe avanzar hacia el modelo acoplado océano-atmósfera.
- La comprensión de la estructura del programa o programas de cómputo secuencial que implementa/n el modelo acoplado océano-atmósfera CSIRO MII. Se deben estudiar las

transformaciones de los datos de entrada, los (sub)programas secuenciales asociados a cada uno de los modelos por separado (oceánico y atmosférico respectivamente) y del programa asociado al modelo acoplado. También será importante la aplicación de programas específicos de análisis ejecución y de rendimiento. Dos de los resultados más importantes de este estudio serían la identificación de los procesos físicos modelados en cada subrutina y la posibilidad de documentación adecuada de los programas secuenciales (que actualmente no cuentan con esta documentación).

3. RESULTADOS OBTENIDOS/ESPERADOS

Ya se ha llevado a cabo un estudio de la evolución del modelo y se ha hecho un primer análisis sencillo del mismo. El código original del modelo CSIRO evolucionó hasta la versión MII, la cual fue desarrollada para supercomputadoras (CRAY, Silicon Graphics, etc.) con varios procesadores bajo sistema operativo UNIX. Para su utilización en la UIOM en el CENPAT fue necesario portar el modelo en computadores de rango intermedio (Sun Ultra10, Enterprise) con un solo procesador. En una etapa posterior se implementó también en la supercomputadora Cray Origin 2000, instalada en las dependencias de la Secretaría de Ciencia, Tecnología e Innovación Productiva (SeCyT) para uso científico [6].

En la actualidad, la ejecución del modelo CSIRO MII se implementa en computadoras personales (PCs) con procesadores Intel, bajo sistema operativo Linux. La versión disponible del código fuente del modelo CSIRO MII, es un claro ejemplo de lo que en la bibliografía se identifica como “código heredado” ó “*legacy code*” [8] [7]. Fue programado en lenguaje Fortran 77, compuesto de un programa principal y aproximadamente 250 subrutinas. No se dispone de documentación sobre el ciclo de vida del software inicial y la documentación incluida en el código es obsoleta, relata aportes y modificaciones, no especifica funcionalidades, ni refleja su estructura. La capacitación recibida oportunamente por especialistas de la UIOM del CENPAT fue orientada a su uso “tal como está”, no a su mantenimiento ni modificación.

Este proyecto puede ser considerado como uno de los avances del presentado en [10] y es resultado directo de que cualquier intento de comprensión del software del modelo CSIRO MII, ya sea para optimizar su rendimiento como para su paralelización, que implica una primera etapa de recuperación del diseño del programa [8]. Para ello se emplea la ingeniería inversa. El objetivo principal de este proyecto es recuperar el diseño del modelo climático acoplado océano-atmósfera (CSIRO MII) en un nivel alto de abstracción, es decir:

- Un diagrama de estructura de la comunicación entre los modelos numéricos involucrados: acoplado, atmosférico y oceánico.
- Para cada modelo (acoplado, atmosférico y oceánico):
 - Árbol de invocación.
 - Listado de las rutinas utilizadas.
 - Listado de bloques *common* de Fortran 77 de entrada.
 - Código re-documentado de las subrutinas.
- Implementación de reglas para automatizar el proceso de depuración del código fuente, que básicamente serían:
 - Quitar líneas de comentario referidas a evolución histórica pero que no aportan información sobre diseño ni funcionalidad.

- Quitar líneas de comentario que anulan código en desuso.

La consecución de estos objetivos permitirá, además, la consolidación del grupo de investigación como tal, avanzar en el conocimiento sobre modelos numéricos y cómputo paralelo, así como incorporar becarios/tesistas de maestría y/o doctorado, con la consecuente formación de recursos humanos.

4. BIBLIOGRAFÍA

- [1] Intergovernmental Panel on Climate Change, Tercer Informe de Evaluación – Cambio Climático 2001 – Base Científica, disponible en <http://www.ipcc.ch/pub/un/ipccwg1s.pdf>
- [2] Welcome to the Commonwealth Scientific and Industrial Research Organisation (CSIRO), <http://www.csiro.au/>
- [3] G. J. Boer, K. Arpe, M. Blackburn, M. Déqué, W. L. Gates, T. L. Hart, H. le Treut, E. Roeckner, D. A. Sheinin, I. Simmonds, R. N. B. Smith, T. Tokioka, R. T. Wetherald, D. Williamson, “Some Results From an Intercomparison of the Climates Simulated by 14 Atmospheric General Circulation Models”, J. Geophys. Res., 97(D12), 12,771–12,786.
- [4] G. J. Boer et al, “Intercomparison of climates simulated by 14 atmospheric general Circulation model”, WMO/TD-No 425, CAS/JSC Working Group on Numerical Experimentation.
- [5] Global and Planetary Change, An Overview of results from the Coupled Model Intercomparison Project (CMIP), Feb. 2001, UCRL-JC 140274.
- [6] <http://www.supercomputo.secyt.gov.ar/>
- [7] Piattini M., J. Villalba, F. Ruiz, R. Bastabchury, M. Polo, M. Martinez, C. Nistal, Mantenimiento del Software-Modelos, técnicas y métodos para la gestión del cambio, Alfaomega Ra-Ma.
- [8] Pressman R., adaptado por Darle Ince, Ingeniería del Software Un enfoque práctico, (quinta edición), Mc Graw Hill, 2002.
- [9] Program for Climate Model Diagnosis and Intercomparison <http://www-pcmdi.llnl.gov/>
- [10] F. G. Tinetti, P. G. Cajaraville, J. C. Labraga, M. A. López, "Cómputo Paralelo Aplicado a Modelos Numéricos del Clima", IX Workshop de Investigadores en Ciencias de la Computación, Universidad Nacional de La Patagonia San Juan Bosco (UNPSJB), Trelew, Chubut, Argentina, Mayo 3-4 de 2007.
- [11] World Meteorological Organization, Numerical Method Used in Atmospheric Models, GARP Publication Series 17, 1979.